Do These Sequences Make CENs Yet?

James A. Birchler

Genome Res. 1997 7: 1035-1037
Access the most recent version at doi:10.1101/gr.7.11.1035

Do These Sequences Make CENs Yet?

James A. Birchler[1]

Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211-7400 USA

Although there is considerable debate about just what sequences constitute a centromere in multicellular eukaryotes, a picture is emerging that degenerate tandem arrays spanning the megabase range are involved. With the general organization of *Arabidopsis* centromeres now known (Round et al., this issue), there are three types of centromeres from plants in addition to human and *Drosophila* for which the physical organization has been studied. A unit repeat present thousands of times at each site is a common theme.

Comparison of Known Centromeric Sequences

The first centromere sequence of a multicellular eukaryote cloned and analyzed was the neocentromere of maize (Peacock et al. 1981). The neocentromeres (or chromosomal knobs) are normally heterochromatic landmarks on several maize chromosomes. There is, in fact, considerable variation for their chromosomal positions. Under normal conditions they are inert, but in the presence of an abnormal chromosome 10, they acquire centromere functions during meiosis (Rhoades and Vilkomerson 1942). These functions supercede those of the normal centromere.

The neocentromere repeat is 185 bp in length and is repeated at one site an estimated 15,000 copies, at a minimum (T. Phelps, pers. comm.). The available evidence suggests that the unit repeat is the sole constituent of neocentromeres, although single- or low-copy interspersions would be difficult to detect.

The sequences required for centromeres in humans are still a matter of discussion, but the alphoid repeat seems to be present in almost all cases (Tyler-Smith et al. 1993; Harrington et al. 1997; Kipling and Warburton 1997). The unit repeat is 171 bp in length. Individual centromeres have clusters of distinguishable alphoid variants and can span stretches of DNA of 2–3 Mb in length (Tyler-Smith et al. 1993).

Analysis of a normal plant centromere from the supernumerary B chromosome of maize (Alfenito and Birchler 1993; Kaszas and Birchler 1996) indicated that this centromere spans ~9 Mb and is composed of a sequence repeat, highly degenerate in length but that has a plurality unit of 1.4 kb. Part of this repeat shows strong homology to a maize neocentromere unit sequence of 90 bp.

The *Drosophila* centromere in minichromosome Dp1178 appears to be the least similar to the others. A very short repeat is present for most of the centromere, but it is interrupted by low-copy sequences that appear to be critical for function (Murphy and Karpen 1995). This difference in centromere organization might explain the difference in isochromosome formation between *Drosophila* and plants: In plants, univalent centromeres, as present in trisomics or monosomics, undergo ''misdivision''—the two arms proceed to opposite poles (Darlington 1939; Sears 1952) resulting in breakage at the centromere. The replicated chromosome arms often fuse at the break in the centromere to produce the mirror image or isochromosome. This phenomenon is found in many species throughout the plant kingdom. Such behavior is consistent with the proposed organization of plant centromeres as repetitive structures that can be subdivided and still function. In contrast, in *Drosophila*, isochromosome formation appears to be caused by translocations between the centric heterochromatin rather than centromere fission, and the central core organization is consistent with this finding.

Centromere sequences isolated from several species among the grasses also show repetitive sequences. This family of repeats likewise has homology to the maize neocentromere unit. In addition, it hybridizes to centromeres of barley, wheat, rye, maize, and rice (Aragon-Alcaide et al. 1996).

There is, however, another sequence that has very strong conservation throughout the grasses but appears to be distinct from the above described sequences (Jiang et al. 1996). This conserved sequence hybridizes to all known grass centromeres including B chromosomes, as determined by in situ hybridization studies, but it does not hybridize to the maize neocentromere. Its strong conservation argues for an important function in the action of grass centromeres, but its lack of any homology to the knob unit repeat suggests that it is not absolutely required for more general centromere function.

Round et al. (this issue) report the organization of the centromeres of *Arabidopsis*. A unit repeat of 180 bp is present in arrays that probably span more than 1 Mb at each centromeric site. Although some interspersions of other sequences were noted, two-dimensional restriction digests confirm the organization as tandem arrays. The arrangement most closely parallels the maize neo and the human centromeres.

Despite evidence that all centromeres in the grasses have homology, there is clearly divergence in function. Crosses of wheat and oat with maize pollen will allow fertilization. However, in the early mitotic divisions of the zygote, the sperm-derived maize chromosomes become lost (Laurie and Bennett 1989; Rines and Dahleen 1990). This loss is correlated with an increasingly diffuse cytological appearance of the maize centromeres. It is as if the maize centromeres cannot organize the centromere proteins following replication in the foreign environment.

Interestingly, in the oat–maize crosses, a low frequency of cases are found in which a single maize chromosome survives and becomes stabilized in transmission from generation to genera-

[1]E-MAIL birchler@biosci.mbp.missouri.edu; FAX (573) 882-0123.

tion (Ananiev et al. 1997). Because different maize chromosomes can be recovered in this manner, there seems to be no DNA sequence property involved. Rather it appears that a random stabilization occurs that can then be perpetuated through the generations. Once the oat and maize centromere organization is elucidated, it will be interesting to investigate this phenomenon.

Analysis of the B chromosome of maize provided a demonstration that the repetitive sequences at plant centromeres are included in the functional kinetochore (Alfenito and Birchler 1993; Kaszas and Birchler 1996). By using successively rearranged chromosomes caused by centromere fission or misdivision, the central portion of the B centromere was displaced to a new location. Because this portion still contained abundant copies of the repeat and the chromosome was transmitted throughout the life cycle and meiosis, the repeat must be a major constituent of the kinetochore. As noted above, this repeat shows strong DNA sequence homology to a portion of the maize neocentromere, where the organization of DNA also appears to be solely a reiteration of a basic repeat. Further subdivision of the B centromere revealed the retention of at least one 55-kb *Eco*RI restriction fragment (of which there are several in the full centromere) in all reduced derivatives, whereas all other fragment sizes could be eliminated without consequence. The retained 55-kb fragments in different derivatives, however, could be distinguished from one another by two-dimensional restriction digests. This result suggests an organization in which the 9-Mb centromere has several conserved regions, any one of which can provide full centromere function and are surrounded by less critical sequences. Nevertheless, the two-dimensional analysis indicated that the basic repeat is still the major constituent of the presumptive critical regions. Further work is required to determine whether any short unique sequences are present at these selected sites.

Inhibition of Recombination by Centromeres

The finding of Round et al. (this issue) that no recombinants were found among centromere variants from any of the five *Arabidopsis* chromosomes but were recovered among the ribosomal RNA repeats in the same mapping lines poses some interesting questions about centromere sequence evolution. In most species, there is a strong inhibition of meiotic recombination surrounding the centromeres. With the elucidation that all centromeres in a species are composed of megabase stretches of homologous sequences, an evolutionary explanation for the centromere reduction of recombination emerges. Clearly, recombination between centromeres of nonhomologous chromosomes would generate reciprocal translocations, which in many species would be associated with partial sterility. In meiotic spreads of maize chromosomes, centromere ''fusions'' composed of two or three nonhomologous chromosomes, are common. The neocentromeres (chromosomal knobs) also form fusions among themselves and, occasionally, with normal centromeres. Despite these quite regular associations, no translocations have ever been observed to occur spontaneously between normal or neocentromeres in direct screens for them (Weber and Alexander 1972). A centromere inhibition of recombination would prevent the formation of such translocations under normal circumstances.

In this vein, it is interesting to reflect on the formation of translocation complexes that have arisen several times in higher plant evolution in unrelated genera such as Oenothera, Paeonia, Isotoma, Rheo, and others. The best studied of these is Oenothera (Cleland 1962). In many species of this genus, all the chromosomes of the genome are involved in centromere translocations. These ''complexes'' are associated with recessive lethals and exist in a reoccurring heterozygous state with normal untranslocated chromosomes. The finding that plant centromeres are long arrays of homologous sequences raises the possibility that the translocations in these species arose from recombination between centromeres of nonhomologous chromosomes. The alternative is that there is an incredibly high rate of random translocation formation together with intense and rapid selection against all other translocations rather than against those between centromeres. All possible combinations of exchanges of chromosome arms have been documented in Oenothera without any other changes in the length of these elements, suggesting that the translocations forming the complexes were restricted to the centromere regions. Nevertheless, crossing-over in Oenothera appears to be restricted to the ends of the chromosomes (Cleland 1962), indicating special circumstances for generating centromere translocations.

The inhibition of recombination within centromere regions and between centromeres of nonhomologous chromosomes raises questions of how the homogenization of centromere sequences within a species can occur. The homogenization of ribosomal RNA repeats at the nucleolar organizing region (NOR) has been proposed to result from unequal crossing-over among the tandem units. A possible solution for homogenization of centromere regions has been suggested by Kipling and Warburton (1997). They note the presence of a CENP-B binding site in the alphoid repeat as well as in certain transposable elements. A consensus binding site for CENP-B also exists in the grass repeat isolated by Aragon-Alcaide et al. (1996), in the maize neocentromere repeat (T. Phelps, pers. comm.), and in the *Arabidopsis* centromere unit. If the CENP-B protein is involved in DNA nicking to promote recombination, the homogenization of repeat arrays could occur. If this is achieved by gene conversion via a mechanism similar to gap repair after P-element transposition in *Drosophila* (Nassif et al. 1994; Thompson-Stewart et al. 1994), homogenization could occur without the generation of reciprocal translocations. The inhibition of meiotic recombination surrounding centromeres would be selected to prevent translocation formation, and the presence of a CENP-B site in centromere repeats would be selected to foster the maintenance of similar sequences at the centromeres of nonhomologous chromosomes. Clearly, further work is required to understand this situation.

Barbara McClintock recognized that recombination between nonhomologous chromosomes could occur in haploid plants (McClintock 1933). Translocations can be recovered from these haploids, but none have ever been observed between centromeres or neocentromeres, illustrating that the lack of recombination between centromeres of nonhomologous chromosomes is not just due to steric hindrance in a diploid. If the *Arabidopsis* centromeres are ~1 Mb

each, then the fraction of a genome devoted to centromeres is ~5%. If this percentage is also typical of maize, crossing-over in the haploid among centromere sequences could be detected were it to occur.

Interestingly, McClintock later described a genetic element termed the X factor, which when present, promoted reciprocal exchange between centromeres, between neocentromeres, or between centromeres and neocentromeres (McClintock 1978). We now know that there is sequence relatedness among these cytological features. The fact that this element promotes these types of rearrangements indicates that they can occur and that under normal conditions they are inhibited from forming. It is not clear whether these rearrangements are generated throughout the life cycle or only meiotically. With the availability of both normal centromere and neocentromere probes, it should now be possible to address the molecular nature of these rearrangements. By defining the junction sequences, aspects of the mechanism could possibly be revealed. Such studies would provide information on the structure and function of centromeres as well as on the induction of exchange between centromeres in the presence of X factor to overcome the normal inhibition.

Artificial Chromosomes

The identification of centromeric sequences in several multicellular eukaryotes has led to the hope that artificial chromosomes can be constructed in the respective species. Artificial chromosome constructs should prove useful in defining centromere functions and chromosomal behavior as well as providing a tool for studies of gene expression and practical applications. In human cells, artificial chromosomes have been assembled in vivo from constituent sequences introduced together (Harrington et al. 1997). Because the size of these selected chromosomes is 6–10 Mb in length, this is probably the optimum range. Fortunately, this size is not the minimum.

Eventually, one would want to design an artificial chromosome to specification, such that single-copy genes of choice are included together with the centromere, replication origins, and telomeres. The sizes of most centromeres preclude using a full-length DNA fragment with present day molecular techniques. However, the divisibility of centromeres because of their repetitive structure provides hope that attempts to produce predesigned artificial chromosomes will soon be possible. The centromere of the maize B chromosome was divided by successive misdivision from a progenitor size of ~9 Mb to derivatives with centromeres of only a few hundred kilobases (Kaszas and Birchler 1996). Because the breaks involved to produce these chromosomes were random and a small sample, it is likely that centromeres in the 100- to 200-kb range might function sufficiently well. This suggests that current techniques could be used to manipulate them.

Futuristically, especially given that it is becoming clear that centromeres from a wide range of species all share the presence of a repeat motif, one might imagine that artificial chromosomes using such derivative centomeres could be devised for the transfer of very large genes for detailed study, gene therapy applications, or the introduction of complete biochemical pathways into agricultural crops for production of selected metabolites.

REFERENCES

Alfenito, M.R. and J.A. Birchler. 1993. *Genetics* 135: 589–597.

Ananiev, E., O. Riera-Lizarazu, H.W. Rines, and R.L. Phillips. 1997. *Proc. Natl. Acad. Sci.* 94: 3524–3529.

Aragon-Alcaide, L., T. Miller, T. Schwarzacher, S. Reader, and G. Moore. 1996. *Chromosoma* 105: 261–268.

Cleland, R.E. 1962. *Adv. Genet.* 11: 147–237.

Darlington, C.D. 1939. *J. Genet.* 37: 341–364.

Harrington, J.J., G. Van Bokkelen, R.W. Mays, K. Gustashaw, and H.F. Willard. 1997. *Nature Genet.* 15: 345–354.

Jiang, J., S. Nasuda, F. Dong, C.W. Scherrer, S.-S. Woo, R.A. Wing, B.S. Gill, and D.C. Ward. 1996. *Proc. Natl. Acad. Sci.* 93: 14210–14213.

Kaszas, E. and J.A. Birchler. 1996. *EMBO J.* 15: 5246–5255.

Kipling, D. and P.E. Warburton. 1997. *Trends Genet.* 13: 141–145.

Laurie, D.A. and M.D. Bennett. 1989. *Genome* 32: 953–961.

McClintock, B. 1933. *Z. Zellforsch. Mikrosk. Anat.* 19: 191–237.

———. 1978. *Stadler Genet. Symp.* 10: 25–47.

Murphy, T.D. and G.H. Karpen. 1995. *Cell* 82: 599–609.

Nassif, N., J. Penney, S. Pal, W.R. Engels, and G.B. Gloor. 1994. *Mol. Cell. Biol.* 14: 1613–1625.

Peacock, W.J., E.S. Dennis, M.M. Rhoades, and A.J. Pryor. 1981. *Proc. Natl. Acad. Sci.* 78: 4490–4494.

Rhoades, M.M. and H. Vilkomerson. 1942. *Proc. Natl. Acad. Sci.* 28: 433–436.

Rines, H.W. and L.S. Dahleen. 1990. *Crop Sci.* 30: 1073–1078.

Round, E.K., S.K. Flowers, and E.J. Richards. 1997. *Genome Res.* (this issue).

Sears, E.R. 1952. *Chromosoma* 4: 535–550.

Thompson-Stewart, D., G.H. Karpen, and A.C. Spradling. 1994. *Proc. Natl. Acad. Sci.* 91: 9042–9046.

Tyler-Smith, C., R.J. Oakey, Z. Larin, R.B. Fisher, M. Crocker, N.A. Affara, M.A. Ferguson-Smith, M. Muenke, O. Zuffardi, and M.A. Jobling. 1993. *Nature Genet.* 5: 368–375.

Weber, D.F. and D.E. Alexander. 1972. *Chromosoma* 39: 27–42.